Stochastic Processes and their Applications 50 (1994) 83–99
North-Holland

Almost sure convergence of a class of stochastic algorithms

Jean Claude Biscarat

LSTA, Université Paris VI, France

Received 24 February 1992
Revised 5 April 1993

In this paper, we establish the almost sure convergence of $\mathbb{R}^d$-valued sequences generated by a particular class of stochastic algorithms and we apply this result to a stochastic approximation type EM algorithm for the mixture problem.

Monte Carlo EM algorithms * stopping times * mixtures of distributions

0. Introduction

The aim of this paper is to investigate the asymptotic behavior of a sequence $\{\phi_n, n\in\mathbb{N}\}$ of $\mathbb{R}^d$-valued r.v.'s generated by an algorithm of the following form:

$$\phi_0 \in G\,,$$
$$\phi_{n+1} = \begin{cases} T(\phi_n) + \gamma_n V(\phi_n, z_n(\phi_n)) & \text{if } T(\phi_n) + \gamma_n V(\phi_n, z_n(\phi_n)) \in G\,, \\ \phi_0 & \text{otherwise}\,, \end{cases} \tag{0.1}$$

where G is a compact subset of $\mathbb{R}^d$, $d\geqslant 1$, T is a C^2 function $\mathbb{R}^d \to \mathbb{R}^d$, $\{\gamma_n, n\in\mathbb{N}\}$ is a sequence of positive constants decreasing to zero, V is a measurable mapping $G\times\mathbb{R}^d \to \mathbb{R}^d$, and for each $\phi\in G$, $\{z_n(\phi), n\in\mathbb{N}\}$ denotes a sequence of independent $\mathbb{R}^d$-valued r.v.'s.

We establish in Theorem 1 that, under suitable conditions, the sequence ϕ_n converges almost surely (a.s.) to a stable fixed point of T. This result is obtained for a general class of stochastic algorithms but the main application concerns stochastic versions of the Expectation Maximization algorithm (EM) (Dempster, Laird and Rubin, 1977), namely the Stochastic Approximation EM algorithm (SAEM) (Celeux and Diebolt, 1991) and the Monte Carlo EM algorithm (MCEM) (Tanner and Wei, 1991). For a monography introducing a review of these algorithms and related topics, see, e.g., Tanner (1991).

On the one hand, the SAEM algorithm introduced by Celeux and Diebolt (1993) is exactly of the form (0.1). In the mixture of distributions problem, Celeux and Diebolt have

Correspondence to: Dr. Jean Claude Biscarat, 121, rue Jeanne d'Arc, 75013 Paris, France.

SSDI 0304-4149(93)E0033-B

proved the almost sure convergence of the SAEM sequence but their proof is crucially depending on the boundedness property of the r.v.'s $V(\phi_n, z_n(\phi_n))$ involved by the mixture context. In Theorem 1 this boundedness property is relaxed and replaced by a weaker condition (see (C11)) on the rate of decay of the tail probabilities of the r.v.'s $V(\phi_n, z_n(\phi_n))$. Then, it allows us to extend the field of application of the SAEM algorithm to incomplete data problems in which the r.v.'s $V(\phi_n, z_n(\phi_n))$ are no longer uniformly bounded, for instance censored data and missing values in multivariate samples.

On the other hand, the MCEM algorithm has been introduced by Wei and Tanner (1991) but no convergence result concerning this algorithm has been established yet. In the second part of this paper, we focus on the mixture of distributions problem and we prove in Proposition 2 that the MCEM sequence can be expressed under the form (0.1). Then, using Theorem 1, we prove the almost sure convergence of this algorithm in Theorem 2.

This paper is organized as follows: In Section 1, after introducing some notations and basic assumptions, we state Theorem 1 as well as relative technical lemmas and we briefly sketch the main steps of the proof of Theorem 1. In Section 2 we give the proofs of our results. Finally, Section 3 is devoted to an application: in the mixture context, we establish the almost sure convergence of the MCEM algorithm.

All the detailed proofs of our results can be found in Biscarat (1992).

1. Results

For simplicity's sake we set $V_n = V(\phi_n, z_n(\phi_n))$. We denote by $B(x, r)$ the open ball with center x and radius r, $B(x, r) = \{x+h, \|h\| < r\}$, where $\|x\| = \langle x, x\rangle^{1/2}$ is the Euclidean norm on $\mathbb{R}^d$, and by $[u]$ the larger integer $\leqslant u$.

A fixed point ϕ of T is said to be stable if and only if all the eigenvalues of the Jacobian matrix $\mathrm{D}T(\phi)$ have modulus smaller than one; unstable if and only if there exists at least one eigenvalue of $\mathrm{D}T(\phi)$ whose modulus is larger than one; indifferent if and only if all the eigenvalues of $\mathrm{D}T(\phi)$ have modulus equal to one. We denote by F, FS, FU and FI the sets of the fixed points, stable fixed points, unstable fixed points and indifferent fixed points of T contained in G, respectively.

Assumptions on T

(C1) For any fixed point ϕ of T, the Jacobian operator $\mathrm{D}T(\phi)$ is diagonalizable and its eigenvalues are positive real numbers.

(C2) For any fixed point ϕ of T, there exists a symmetric definite positive matrix A (depending on ϕ) such that

$$\langle \mathrm{D}T(\phi)\cdot u, v\rangle_A = \langle u, \mathrm{D}T(\phi)\cdot v\rangle_A \quad \text{for all } u, v \in \mathbb{R}^d,$$

where $\langle\cdot,\cdot\rangle_A$ denotes the inner product defined by $\langle u, v\rangle_A = \langle A\cdot u, v\rangle$.

(C3) There exists a positive real number r such that $B(T(\phi), r) \subset G$ for all $\phi \in G$.

(C4) The set F is finite, FS is non-empty and FI is empty.

Furthermore, we assume that there exists a C^2 real-valued function L defined on G such that:

(C5) For any fixed point ϕ of T, the matrix $D^2(LT-L)(\phi)$ is definite positive, where LT denotes $L \circ T$.

(C6) For any $\phi \in G$ such that $T(\phi) \neq \phi$ we have $L(T(\phi)) > T(\phi)$.

(C7) For any fixed point ϕ of T, the Jacobian operator $DL(\phi) = 0$ (i.e. ϕ is a stationary point of L).

Remark 1.1. In the mixture context, if T is the operator of the EM algorithm, conditions (C1), (C2), (C6) and (C7) are satisfied and (C5) can be replaced by the following weaker condition: the operator $D^2L(\phi)$ is regular (see Redner and Walker, 1984, and Celeux and Diebolt, 1991). In a forthcoming work, we will prove that, under a stronger condition than condition (C11) below, (C4) can be partially relaxed.

Assumptions on the sequence γ_n

Letting $r_n = n^{\delta_0}$, where δ_0 is a suitable positive constant defined in (2.2):

(C8) The sequence $\{\gamma_n r_n, n \in \mathbb{N}\} \downarrow 0$ and $\sum_{n=0}^{\infty} \gamma_n r_n = \infty$.

(C9) There exists an integer $d_0 \geqslant 2$ such that $\gamma_n r_n^{2d_0} = o(1)$ and $(\gamma_n r_n^{2d_0})^{-1} = o(r_n)$.

$$\text{(C10)} \quad \lim_{n \to \infty} \frac{\gamma_n r_n}{\gamma_{n+1} r_{n+1}} = 1 .$$

Remark 1.2. It can be checked that, if γ_n is of the form $n^{-\theta}$ for some $0 < \theta < 1 + \delta_0$, then the assumptions (C8), (C9) and (C10) are satisfied.

Assumptions on the sequence V_n

$$\text{(C11)} \quad \sum_{n=0}^{\infty} \sup_{\phi \in G} P(\|V(\phi, z_n(\phi))\| \geqslant k r_n) < \infty \quad \text{for all } k > 0 .$$

(C12) For any $\phi^* \in \mathrm{FR}$, there exist two positive constants a and ρ and an integer n_0 such that, for all $n > n_0$,

$$\inf_{\phi \in G} P\{\langle V(\phi, z_n(\phi)), u\rangle_A > a\} > \rho, \qquad \inf_{\phi \in G} P\{\langle V(\phi, z_n(\phi)), u\rangle_A < -a\} > \rho,$$

where $A = A(\phi^*)$ has been introduced in (C2) and u satisfies $\|u\|_A = 1$ and $DT(\phi^*) \cdot u = \lambda u$ for some $\lambda > 1$.

Before stating Theorem 1, we prove in Proposition 1 below that, after a finite number of iterations, the algorithms remains almost surely in G. Thus, in the sequel, we will consider, without loss of generality, that ϕ_n always lies in G.

Proposition 1. *There exists a.s. a finite stopping time N such that, for all $n > N$, the event $\{T(\phi_n) + \gamma_n V_n \in G\}$ occurs.*

Proof. By (C3),

$$P(\{T(\phi_n) + \gamma_n V_n \notin G\}) \leqslant \sup_{\phi \in G} P(\{\|V(\phi, z_n(\phi))\| \geqslant r/\gamma_n\}) .$$

Thus, combining (C8), (C11) and the Borel–Cantelli lemma completes the proof. □

Theorem 1. *For any $\phi_0 \in G$ the sequence $\{\phi_n, n \in \mathbb{N}\}$ defined by (0.1) converges a.s. to a stable fixed point of T.*

For simplicity's sake, we will make use of the following notations:

$$\mathscr{F}_n = \bigcup_{\phi \in \mathrm{F}} B(\phi, \sqrt{\gamma_n r_n}) , \quad \mathscr{F}\mathscr{U}_n = \bigcup_{\phi \in \mathrm{FU}} B(\phi, \sqrt{\gamma_n r_n}) ,$$

$$\mathscr{F}\mathscr{S}_n = \bigcup_{\phi \in \mathrm{FS}} B(\phi, \sqrt{\gamma_n r_n}) ; \tag{1.1}$$

$$\mathscr{C}_n(\phi) = G \backslash \{x \in \mathbb{R}^d \text{ such that } |\langle x - \phi, u\rangle_A| < \gamma_n r_n^{do}\} , \tag{1.2}$$

where u is a unit vector such that $DT(\phi) \cdot u = \lambda u$ for some $\lambda > 1$;

$$\mathscr{A}_n(H, \phi) = \{\phi_n \notin B(\phi, H\sqrt{\gamma_n r_n})\} \cap \left\{ \bigcup_{p=1}^{\infty} (\phi_{n+p} \in B(\phi, \sqrt{\gamma_{n+p} r_{n+p}})) \right\} . \tag{1.3}$$

The proof of Theorem 1 is organized as follows:

Step 1. We establish that ϕ_n a.s. remains in the sets $G \backslash \mathscr{F}\mathscr{U}_n$ after a finite time. This first step is divided into three parts:

Part 1. We prove in Lemma 1 that, for each $\phi \in \mathrm{FU}$, ϕ_n a.s. visits the sets $G \backslash B(\phi, \gamma_n r_n)$ infinitely often (i.o.).

Part 2. We prove in Lemmas 2 and 4 that, for each $\phi \in \mathrm{FU}$, ϕ_n a.s. visits the sets $\mathrm{C}_n(\phi)$ i.o. Then, from (C9), we deduce that for each $\phi \in \mathrm{FU}$ and $H > 1$, ϕ_n a.s. visits the sets $G \backslash B(\phi, H\sqrt{\gamma_n r_n})$ i.o.

Part 3. We prove in Lemmas 6, 7 and 8 that, for each $\phi \in \mathrm{FU}$, there exists a suitable constant H such that the events $\mathscr{A}_n(H, \phi)$ occurs a.s. for a finite number of n's at most.

Note that in Step 1 we will make use of a crucial but technical result postponed to the Appendix (Corollary A) in the interest of clarity.

Step 2. We establish in Lemma 9 that ϕ_n a.s. visits the sets $\mathscr{F}_n$ i.o. Then, from the first step, we deduce that ϕ_n a.s. visits the sets $\mathscr{F}\mathscr{S}_n$ i.o.

Step 3. We establish in Lemma 10 that, for each $\phi \in \mathrm{FS}$, if ϕ_n enters $B(\phi, \sqrt{\gamma_n r_n})$ for some n sufficiently large, then it remains there a.s. Finally, collecting the results obtained in Steps 1, 2 and 3, we complete the proof of Theorem 1.

Before stating these lemmas we need to introduce some additional notation:

$$q_n = \langle \phi_n - \phi^*, u\rangle_A \quad \text{and} \quad w_n = \langle V_n, u\rangle_A , \tag{1.4}$$

where $\phi^* \in \mathrm{F}$ and u is unit eigenvector of $DT(\phi^*)$;

$$D_n^p = \{ |q_n| > \gamma_n r_n^p \} \quad \text{and} \quad C_n^p = \{ |q_n| < \gamma_n r_n^p \} , \tag{1.5}$$

where p is a positive integer;

$$G_n = G \setminus \bigcup_{\phi \in \mathrm{F}} B(\phi, \sqrt{\gamma_n r_n}) . \tag{1.6}$$

Lemma 1. *For any unstable fixed point ϕ^* of T, there exists a.s. a finite stopping time n_1 such that, if*

$$\tau_1 = \inf\{k > n_1 :\ |q_k| < \gamma_k r_k\} \quad \textit{and} \quad \xi_1 = \inf\{k > \tau_1 :\ |q_k| > \gamma_k r_k\} ,$$

then

$$\{\tau_1 < \infty\} \subset \{\xi_1 < \infty\} .$$

Lemma 2. *For any unstable fixed point ϕ^* of T, there exists a.s. a finite stopping time n_2 such that for all integers $p \leqslant d_0$ and $n > n_2$ we have $D_n^p \cap C_n^{p+1} \subset D_{n+1}^p$.*

Lemma 3. *Let $\{x_n, n \in \mathbb{N}\}$ be a sequence of constants satisfying the following condition:*

$$\sum_{n=0}^{\infty} \sup_{\phi \in G} P\left(\|V(\phi, z_n(\phi))\| \geqslant k \frac{x_n}{\gamma_n} \right) < \infty \quad \textit{for all } k > 0 . \tag{1.7}$$

Then, there exists a constant $K > 0$ such that for any fixed point ϕ^* of T, there exists a.s. a finite stopping time n_3 such that for all integer $n > n_3$,

$$\{ |\phi_n - \phi^*| < x_n \} \subset \{ |\phi_{n+1} - \phi^*| < K x_n \} .$$

Lemma 4. *For any unstable fixed point ϕ^* of T, there exists a.s. a finite stopping time n_4 such that for all integer $p \leqslant d_0$ and $n > n_4$, if*

$$\tau_4 = \inf\{k > n :\ \gamma_k r_k^p < |q_k| < \gamma_k r_k^{p+1}\}$$

and

$$\xi_4 = \inf\{k > \tau_4 :\ |q_k| \notin \,]\gamma_k r_k^p, \gamma_k r_k^{p+1}[\,\} ,$$

then

$$\{\tau_4 < \infty\} \subset \{\xi_4 < \infty\} .$$

Lemma 5. *There exist a.s. a finite stopping time n_5 and a real number $\beta > 0$ such that for all integer $n > n_5$ we have*

$$\{\phi_n \in G_n\} \subset \{L(\phi_{n+1}) - L(\phi_n) \geqslant \beta \gamma_n r_n\} .$$

Lemma 6. *For any unstable fixed point ϕ^* of T, there exist a.s. a finite stopping time n_6 and an integer p_0 such that for all $n > n_6$ if*

$$\tau_6 = \inf\{k > n:\ \phi_k \notin B(\phi^*, \sqrt{\gamma_k r_k})\}$$

then

$$\{\tau_6 < \infty\} \cap \left\{\bigcap_{j=0}^{p_0} \{\phi_{\tau_6+j} \in \mathrm{G}_{\tau_6+j}\}\right\}$$

$$\subset \left\{L(\phi_{\tau_6+p_0}) > \sup_{\phi \in B(\phi^*, \sqrt{\gamma_{\tau_6+p_0} r_{\tau_6+p_0}})} L(\phi)\right\}.$$

Lemma 7. *For any unstable fixed point ϕ^* of T, there exist a.s. a finite stopping time n_7 and a real number $H > 1$ such that for all $n > n_7$, if we define*

$$\tau_7 = \inf\{k > n:\ \phi_k \notin B(\phi^*, \sqrt{\gamma_k r_k}) \quad \textit{and} \quad \phi_{k-1} \in B(\phi^*, \sqrt{\gamma_{k-1} r_{k-1}})\}\,,$$

then

$$\{\tau_7 < \infty\} \subset \{\phi_{\tau_7+p_0} \in B(\phi^*, H\sqrt{\gamma_{\tau_7+p_0} r_{\tau_7+p_0}})\}\,.$$

Lemma 8. *For any unstable fixed point ϕ^* of T, there exists a.s. a finite stopping time n_8 such that for all $n > n_8$, if we define*

$$\tau_8 = \inf\{k > n:\ \phi_k \notin B(\phi^*, \sqrt{\gamma_k r_k});\ \phi_{k-1} \in B(\phi^*, \sqrt{\gamma_{k-1} r_{k-1}})\}$$

and

$$\xi_8 = \inf\{k > \tau_8:\ \phi_k \notin B(\phi^*, H\sqrt{\gamma_k r_k})\}\,,$$

then

$$\{\xi_8 < \infty\} \cap \left\{\bigcap_{j=\tau_8+1}^{\xi_8} \{\phi_j \in G_j\}\right\} \subset \bigcap_{m > \tau_8} \{\phi_m \notin B(\phi^*, \sqrt{\gamma_{\tau_8} r_{\tau_8}})\}\,.$$

Lemma 9. *There exists a.s. a finite stopping time n_9 such that, for all integer $m > n_9$, there exists an integer $n > m$ such that $\phi_n \notin G_n$.*

Lemma 10. *Let $\{\chi_n, n \in \mathbb{N}\}$ be a sequence of constants such that*

$$\sum_{n=0}^{\infty} \sup_{\phi \in G} P\left(\|V(\phi, z_n(\phi))\| \geqslant k\,\frac{\chi_n}{\gamma_n}\right) < \infty \quad \textit{for all } k > 0\,, \tag{1.8}$$

$$\lim_{n\to\infty} \frac{\chi_{n+1}}{\chi_n} = 1\,.$$

For any stable fixed point ϕ^* of T, there exists a norm N^* on $\mathbb{R}^d$ such that for any $c>0$, there exists a.s. a finite stopping time n_{10} such that for all $n>n_{10}$ we have

$$\{N^*(\phi_n-\phi^*)<cy_n\}\subset\{N^*(\phi_{n+1}-\phi^*)<cy_{n+1}\}\ .$$

2. Proofs of the results

Proof of Theorem 1. We start by proving the following result.

Result 1. *With probability one, the sequence ϕ_n ultimately does not visit $\mathscr{F}\mathscr{U}_n$.*

Proof. Since FU is finite, the proof of Result 1 will be completed if we prove that a.s. for each $\phi^*\in$ FU, ϕ_n ultimately does not visit $B(\phi^*, \sqrt{\gamma_n r_n})$. Therefore, in view of Lemma 8, it suffices to prove that for each $\phi^*\in$ FU, if ϕ_n enters $B(\phi^*, \sqrt{\gamma_n r_n})$ for some n sufficiently large, then a.s. it escapes from $B(\phi^*, H\sqrt{\gamma_n r_n})$ after a finite time, whatever the constant $H>1$. Assume that ϕ_n enters $B(\phi^*, \sqrt{\gamma_n r_n})$ at the time τ with $\tau>n_i$ for $1\leqslant i\leqslant 10$, where n_i has been defined in Lemma i, $1\leqslant i\leqslant 10$. We have to distinguish two different cases:

Case 1: $\phi_\tau\in\{|q_\tau|<\gamma_\tau r_\tau\}$. Then, by Lemma 1, there exists a.s. a finite $\xi>\tau$ such that $\phi_\xi\notin\{|q_\xi|\geqslant\gamma_\xi r_\xi\}$.

If $\phi_\xi\notin B(\phi^*, H\sqrt{\gamma_\xi r_\xi})$, then Result 1 is proved.

If $\phi_\xi\in B(\phi^*, H\sqrt{\gamma_\xi r_\xi})$, then, in view of (C9) and (1.5), there exists an integer $p\leqslant d_0$ such that $\phi_\xi\in D_\xi^p\cap C_\xi^{p+1}$. But ϕ_n escapes a.s. from C_n^{p+1} after a finite time: Indeed, if $\phi_n\in\bigcap_{n>\xi}C_n^{p+1}$ then, by Lemma 2, we would have $\phi_n\in\bigcap_{n>\xi}\{D_n^p\cap C_n^{p+1}\}$, which by Lemma 4 is impossible. Hence, using the same arguments as above we obtain that $\phi_{\tau+n}$ escapes a.s. from $\bigcup_{m=1}^{d_0}\{D_{\tau+n}^m\cap C_{\tau+n}^{m+1}\}$ at a finite n, and since by (C9), $\gamma_k r_k^{d_0+1}>H\sqrt{\gamma_k r_k}$ for k large enough, the proof of Case 1 is completed.

Case 2: $\phi_\tau\notin\{|q_\tau|<\gamma_\tau r_\tau\}$. Then, in view of (C9) and (1.5), $\phi_\tau\in\bigcup_{m=1}^{d_0}\{D_\tau^m\cap C_\tau^{m+1}\}$, so we are again in the same situation as in Case 1 above, consequently the proof of Result 1 is completed. □

Now let us return to the proof of Theorem 1: Result 1 asserts that there exists a.s. a finite θ such that, for all $n>\theta$, $\phi_n\notin\mathscr{F}\mathscr{U}_n$. By Lemma 10, for any $\phi^*\in$ FS and any $c>0$, there exists a.s. a finite $n_{10}(\phi^*)$ such that, for all $n>n_{10}(\phi^*)$,

$$\{N^*(\phi_n-\phi^*)<c\sqrt{\gamma_n r_n}\}\subset\bigcap_{k>n}\{N^*(\phi_k-\phi^*)<c\sqrt{\gamma_k r_k}\}\ .$$

Now, let $\hat{n}=\max(\theta, n_9, \max_{\phi^*\in\mathrm{FS}}n_{10}(\phi^*))$. Notice that, since FS is finite, $\hat{n}$ is a.s. finite. Moreover, Lemma 9 and Result 1 entail that there exists $n>\hat{n}$ such that $\phi_n\notin\mathscr{F}\mathscr{U}_n\cap G_n$. Consequently, there exists $\phi^*\in$ FS such that $\phi_n\in B(\phi^*, \sqrt{\gamma_n r_n})$. Since all the norms on $\mathbb{R}^d$ are equivalent, it follows that $N^*(\phi_n-\phi^*)<b_*\sqrt{\gamma_n r_n}$ for some positive constant b_*. Finally by Lemma 10 we obtain $N^*(\phi_k-\phi^*)<b_*\sqrt{\gamma_k r_k}$ for all $k>n$. So Theorem 1 is established. □

Proof of Lemma 1. First, we need to introduce the following notation:

$$j_n = [b \log n] \quad \text{and} \quad \tilde{n} = n - 1 + \sum_{k=1}^{n-1} j_k , \tag{2.1}$$

where b is a suitable positive constant defined in the Appendix. Also, $S(\phi)$ denotes the set of the eigenvalues of $DT(\phi)$ larger than one, and $\lambda_0 = \min_{\phi \in F} S(\phi)$. The real number δ_0 such that $r_n = n^{\delta_0}$, satisfies the following relation:

$$0 < \delta_0 < b \log \lambda_0 . \tag{2.2}$$

Consider an unstable fixed point, ϕ^*, of T and an eigenvector, u, of $DT(\phi^*)$ such that $DT(\phi^*) \cdot u = \lambda \cdot u$ with $\lambda > 1$. By (C2),

$$q_{n+1} = q_n + O(q_n^2) + \gamma_n w_n , \tag{2.3}$$

from which it follows that

$$q_{\tilde{n}+j_n} = \lambda^{j_n} q_{\tilde{n}} + \sum_{k=0}^{j_n - 1} \lambda^k [O(q^2_{\tilde{n}+j_n-k-1}) + \gamma_{\tilde{n}+j_n-k-1} w_{\tilde{n}+j_n-k-1}] . \tag{2.4}$$

Now, set $\Omega_t = \{ |q_n| < \gamma_n r_n \text{ for all } n > t\}$. It suffices to prove that for any $t > 0$, $P(\Omega_t) = 0$. We proceed by contradiction. Suppose that there exists $t > 0$ such that $P(\Omega_t) \neq 0$. Then, from Corollary A in the Appendix, $P(\Omega_t \cap \{\limsup_{n\to\infty} \{E_{n,j_n}\}\} \neq 0$. For each $\omega \in \Omega_t \cap \{\limsup_{n\to\infty}\{E_{n,j_n}\}\}$, there exists $n > 0$ such that $\omega \in \Omega_t \cap \{E_{n,j_n}\}$. From (2.4) and (C10) it follows that $|q_{\tilde{n}+j_n}(\omega)| \geq \frac{1}{2} a \lambda^{j_n - 1} \gamma_{\tilde{n}+j_n}$ which, by (2.2) and (C10), implies that $|q_{\tilde{n}+j_n}(\omega)| \geq \gamma_{\tilde{n}+j_n} r_{\tilde{n}+j_n}$ for all n large enough. Therefore, we obtain a contradiction with the assumption that ω is in Ω_t. □

Proof of Lemma 2. Consider an unstable fixed point, ϕ^*, of T and an eigenvector, u, of $DT(\phi^*)$ such that $DT(\phi^*) \cdot u = \lambda u$ with $\lambda > 1$. By (2.3) and (C9) there exists $\eta > 0$ such that, for all integers $p \leq d_0$ and n large enough,

$$P(D_n^p \cap C_n^{p+1} \cap \{D_{n+1}^p\}^c) \leq \sup_{\phi \in G} P(\|V(\phi, z_n(\phi))\| > \eta r_n^p) .$$

The Borel–Cantelli lemma and (C11) entail that, for any $p \leq d_0$, there exists a.s. a finite $N(p)$ such that for all $n > N(p)$, $D_n^p \cap C_n^{p+1} \subset D_{n+1}^p$. Finally, taking $n_2 = \max_{p \leq d_0}\{N(p)\}$ completes the proof. □

Proof of Lemma 3. Consider a fixed point, ϕ^*, of T and a sequence, x_n, satisfying (1.7). Since $DT(\phi)$ is continuous, there exist positive constants K and η such that

$$P(\{\|\phi_n - \phi^*\| < x_n\} \cap \{\|\phi_{n+1} - \phi^*\| \geq K x_n\})$$
$$\leq \sup_{\phi \in G} P\left(\left\{\|V(\phi, z_n(\phi))\| \geq \eta \frac{x_n}{\gamma_n}\right\}\right).$$

Then, using the Borel–Cantelli lemma and (1.7), completes the proof. □

Proof of Lemma 4. This proof is similar to that of Lemma 1 and is omitted here. □

Proof of Lemma 5. From Celeux and Diebolt (1991, Lemma 1) and (C5) there exists a constant $\alpha > 0$ such that we have, for n sufficiently large, that

$$\inf_{\phi \in G_n} \{(LT - L)(\phi)\} \geqslant \alpha \gamma_n r_n ,$$

which, using a quadratic Taylor expansion of L about $T(\phi_n)$, implies that

$$\begin{aligned}\{\phi_n \in G_n\} \subset \{L(\phi_{n+1}) - L(\phi_n) \geqslant \alpha \gamma_n r_n \\ + \gamma_n DL(T(\phi_n) + t_n \gamma_n V_n) \cdot V_n\} ,\end{aligned} \tag{2.5}$$

where $t_n \in]0, 1[$ and n is large enough. Moreover, if β is a constant such that $0 < \beta < \alpha$,

$$\begin{aligned}&P(\alpha \gamma_n r_n + \gamma_n DL(T(\phi_n) + t_n \gamma_n V_n) \cdot V_n < \beta \gamma_n r_n) \\ &\quad \leqslant \sup_{\phi \in G} P(\| V(\phi, z_n(\phi)) \| > \psi r_n) ,\end{aligned}$$

where ψ is a positive constant. Then, by (C11), the Borel–Cantelli lemma and (2.5) together complete the proof. □

Proof of Lemma 6. Let ϕ^* be a fixed point of T. For simplicity's sake, throughout this proof we will make use of the following notation:

$$B = B(\phi^*, \sqrt{\gamma_{\tau_6 + p_0} r_{\tau_6 + p_0}}) , \qquad \mathscr{B} = B(\phi^*, K\sqrt{\gamma_{\tau_6} r_{\tau_6}}) ,$$

$$\mathscr{E}_{\tau_6} = \{\tau_6 < \infty\} \cap \left\{ \bigcup_{j=0}^{p_0} \{\phi_{\tau_6 + j} \in G_{\tau_6 + j}\} \right\} .$$

It is enough to proof that there exists an integer $p_0 > 0$ such that

$$\mathscr{E}_{\tau_6} \subset \left\{ L(\phi_{\tau_6 + p_0}) - L(\phi^*) > \sup_{\phi \in B} \{L(\phi)\} - L(\phi^*) \right\} .$$

But, by Lemma 3, it suffices to prove that there exists $p_0 > 0$ such that

$$\begin{aligned}\mathscr{E}_{\tau_6} \subset \Big\{ & L(\phi_{\tau_6 + p_0}) - L(\phi_{\tau_6}) \\ & > \sup_{\phi \in B} |L(\phi) - L(\phi^*)| + \sup_{\phi \in \mathscr{B}} |L(\phi) - L(\phi^*)| \Big\} .\end{aligned}$$

The proof then follows from (C7), (C10) and Lemma 5. □

Proof of Lemma 7. This result is a direct consequence of Lemma 3 and (C10). □

Proof of Lemma 8. Throughout this proof, we denote

$$\mathscr{M} = \{\xi_8 < \infty\} \cap \left\{ \bigcap_{j=\tau_8+1}^{\xi_8} \{\phi_j \in G_j\} \right\}.$$

By Lemmas 6 and 7, there exists $p_0 > 0$ such that $\tau_8 + p_0 < \xi_8$ and

$$\mathscr{M} \subset \left\{ L(\phi_{\tau_8+p_0}) > \sup_{\phi \in B(\phi^*, \sqrt{\gamma_{\tau_8} r_{\tau_8}})} \{L(\phi)\} \right\}.$$

Consequently, by Lemma 5 the sequence $\{\phi_{\xi_8+j}, j \in \mathbb{N}\}$ cannot enter $B(\phi^*, \sqrt{\gamma_{\tau_8} r_{\tau_8}})$ before visiting $\bigcup_{\phi' \in \mathrm{F}, \phi' \neq \phi^*} B(\phi', \sqrt{\gamma_{\xi_8+j} r_{\xi_8+j}})$. But, since F is finite we have $\mathscr{M} \subset \{\bigcap_{j=\xi_8+1}^{\xi_8+l} \{\phi_j \in G_j\}\}$ for each l and ξ_8 sufficiently large. Moreover, (C7) and Lemma 5 imply that l can be chosen such that

$$\mathscr{M} \subset \left\{ L(\phi_{\xi_8+l}) - L(\phi_{\xi_8}) > 2 \sup_{\phi^* \in \mathrm{F}} \left(\sup_{\phi \in B(\phi^*, \sqrt{\gamma_{\xi_8} r_{\xi_8}})} |L(\phi) - L(\phi^*)| \right) \right\},$$

which entails that

$$\mathscr{M} \subset \bigcap_{j>0} \left\{ L(\phi_{\xi_8+j}) > \sup_{\phi \in B(\phi^*, \sqrt{\gamma_{\tau_8} r_{\tau_8}})} L(\phi) \right\}. \quad \square$$

Proof of Lemma 9. Let $t > 0$ be such that $P(n_9 < t) \neq 0$. Suppose that $P(\bigcap_{k>t} \{\phi_k \in G_k\}) \neq 0$ and consider $\omega \in \bigcap_{k>t} \{\phi_k \in G_k\}$. From Lemma 5 and (C8),

$$\lim_{n \to \infty} L(\phi_n(\omega)) \geqslant \inf_{\phi \in G} \{L(\phi)\} + \beta \lim_{n \to \infty} \left(\sum_{k=t+1}^{n} \gamma_k r_k \right) = \infty,$$

which contradicts the boundedness of L on G. $\square$

Proof of Lemma 10. Let χ_n be a sequence satisfying (1.8) and consider a stable fixed point, ϕ^*, of T and the spectral radius, λ^*, of $DT(\phi^*)$ and let $\eta > 0$ be such that $\lambda^* + \eta < 1$. From Ciarlet (1985), there exists a matricial norm $\mathscr{N}^*$, depending on ϕ^* and η, subordinated to a norm N^* on $\mathbb{R}^d$, such that $\mathscr{N}^*(DT(\phi^*)) < \lambda^* + \eta$. Let J_n denote the event $\{N^*(\phi_n - \phi^*) < c\chi_n\}$, where $c > 0$. Then, we have

$$P(J_n \cap J_{n+1}^c) \leqslant \sup_{\phi \in G} P\left(\left\{ \|V(\phi, z_n(\phi))\| \geqslant \sigma \frac{\chi_n}{\gamma_n} \right\} \right) \quad \text{for some } \sigma > 0.$$

Finally, by (1.8), an application of the Borel–Cantelli lemma completes the proof. $\square$

3. Applications

In the mixture of distributions problem, Biscarat, Celeux and Diebolt (1991) introduced in a detailed way the simulated annealing version of the MCEM algorithm of Wei and Tanner

(1991). In this section, we first briefly recall their results in order to prove that this algorithm can take the form (0.1). Then, we check that the assumptions ensuring the validity of Theorem 1 are satisfied, which enables us to establish the almost sure convergence of the MCEM sequence.

The observed $\mathbb{R}^m$-valued sample $\boldsymbol{x}=(x_1,\ldots,x_N)$ is assumed to be drawn from the mixture density

$$h(x)=\sum_{k=1}^{K} p^k h(x, a^k) ,$$

where the mixing weights p^k satisfy $0<p^k<1$ and sum to one and the densities $h(x, a^k)$ are distinct members of the same exponential family: The generic density $h(x, a)$ has the form

$$h(x, a)=D(a)^{-1}\tau(x)\exp(a^{\mathrm{T}}b(x)) ,$$

where a is a vector of $\mathbb{R}^S$, a^{T} denotes the transpose of a and $\tau:\mathbb{R}^r\to\mathbb{R}$ and $b:\mathbb{R}^r\to\mathbb{R}^S$ are functions. We have to estimate the parameter $\phi=(p^1,\ldots, p^K, a^1,\ldots, a^K)\in\mathbb{R}^d$, where $d=K+SK$.

First we describe the incomplete data structure of the problem. Let $\boldsymbol{y}=(\boldsymbol{x}, \boldsymbol{z})=\{(x_i, z_i), i=1,\ldots, N\}$ denote the complete data, where the vector of indicator variables $z_i=(z_{ij}, j=1,\ldots, K)$ is defined by $z_{ij}=1$ or 0 according as whether x_i has been drawn from the density $h(x, a^j)$ or not. The r.v.'s $z_1,\ldots, z_N$ are i.i.d. following a multinomial distribution consisting of one draw from K categories with probabilities $p^1,\ldots, p^K$ respectively.

Suppose that $\boldsymbol{y}$ has been generated from the density $g(\boldsymbol{y}, \phi)$ and let $k(\boldsymbol{z}/\boldsymbol{x};\phi)$ be the conditional density of $\boldsymbol{z}$ given $\boldsymbol{x}$.

The EM algorithm is directed at finding the global maximizer, or at least a local maximizer of the likelihood function (l.f.) L of the observed data $\boldsymbol{x}$. The EM method replaces the maximization of the unknown l.f. $g(\boldsymbol{y}, \phi)$ of the complete data by successive maximizations of the conditional expectation of $\log g(\boldsymbol{y}, \phi')$ given $\boldsymbol{x}$ for the current fit ϕ_n of the parameter.

Let $Q(\phi, \phi')$ denote the conditional expectation of $\log g(\boldsymbol{y}, \phi')$ given $\boldsymbol{x}$ for the value ϕ of the parameter, i.e. $Q(\phi, \phi')=E(\log g(\boldsymbol{y}, \phi')/\boldsymbol{x};\phi)$. We have in the mixture setup,

$$Q(\phi, \phi')=\sum_{i=1}^{N}\sum_{j=1}^{K} t^j(x_i)(\log p'^j+\log h(x_i, a'^j)) ,$$

where $t_j(x_i)=k(z_i/x_i;\phi)$ if $z_{ij}=1$ is the posterior probability that x_i has been generated from the jth component.

Starting from an initial position ϕ_0, the nth iteration $\phi_{n+1}=T(\phi_n)$ of EM can be summarized as follows (see, e.g., Titterington, Smith and Makov, 1985):

E step: Compute $Q(\phi, \phi_n)$. This reduces to computing the posterior probability

$$t^j_n(x_i) \quad \text{for } i=1,\ldots, N \text{ and } j=1,\ldots, K .$$

$$t^j_n(x_i)=\frac{p^j_n h(x_i, a^j_n)}{\sum_{s=1}^{K} p^s_n h(x_i, a^j_n)} . \tag{3.1}$$

M step: Choose ϕ_{n+1} to maximize $Q(\phi_n, \phi)$ in ϕ, which provides:

$$p^j_{n+1} = \frac{\sum_{i=1}^N t^j_n(x_i)}{N} \quad \text{for } j = 1,\ldots, K, \tag{3.2}$$

and

$$a^j_{n+1} = \frac{\sum_{i=1}^N t^j_n(x_i) b(x_i)}{\sum_{i=1}^N t^j_n(x_i)} \quad \text{for } j = 1,\ldots, K. \tag{3.3}$$

The MCEM algorithm is obtained by incorporating a Monte Carlo (MC) step between the E and the M steps. More precisely, starting from the initial value ϕ_0, the nth iteration $\phi_n \to \phi_{n+1}$ of MCEM can be described as follows:

E step: Compute $t^j_n(x_i)$, $i = 1,\ldots, N$ and $j = 1,\ldots, K$, as in (3.1).

MC step: For $i = 1,\ldots, N$, draw a sequence $\{e_r(x_i, \phi_n),\ r = 1,\ldots, \mu_n\}$ of i.i.d. random indicator variables $e_r(x_i, \phi_n) = (e^1_r(x_i, \phi_n),\ldots, e^K_r(x_i, \phi_n))$ from a multinomial distribution with parameters $t^1_n(x_i),\ldots, t^K_n(x_i)$, where μ_n is a sequence of integers such that $\mu_n \to \infty$ as $n \to \infty$. If

$$\frac{\sum_{i=1}^N e^j_r(x_i, \phi_n)}{N} \geqslant c(N) \quad \text{for all } r = 1,\ldots, \mu_n \text{ and all } j = 1,\ldots, K, \tag{3.4}$$

where $c(N)$ is a suitable constant satisfying $0 < c(N) < \frac{1}{2}$, then go to the M step below. If $N^{-1}\sum_{i=1}^N e^j_r(x_i, \phi_n) < c(N)$ for some $r = 1,\ldots, \mu_n$ and some $j = 1,\ldots, K$, then draw the new variables $e^j_r(x_i, \phi_n)$ from some preassigned distribution such that condition (3.4) holds.

M step: Compute ϕ_{n+1} as follows:

$$p^j_{n+1} = \frac{\sum_{i=1}^N \{\mu_n^{-1} \sum_{r=1}^{\mu_n} e^j_r(x_i, \phi_n)\}}{N} \quad \text{for } j = 1,\ldots, K, \tag{3.5}$$

$$a^j_{n+1} = \frac{\sum_{i=1}^N \{\mu_n^{-1} \sum_{r=1}^{\mu_n} e^j_r(x_i, \phi_n)\} b(x_i)}{\sum_{i=1}^N \{\mu_n^{-1} \sum_{r=1}^{\mu_n} e^j_r(x_i, \phi_n)\}} \quad \text{for } j = 1,\ldots, K. \tag{3.6}$$

Before stating the main results of this section, we need the following notation where the integers j and r run in $\{1,\ldots, K\}$ and in $\{1,\ldots, \mu_n\}$, respectively.

$T^j(\phi_n) = N^{-}\sum_{i=1}^N t^j_n(x_i)$ denotes the jth component of $T(\phi_n)$.

$T^{K+j}(\phi_n) = N^{-1}\sum_{i=1}^N t^j_n(x_i) b_j(x_i)$ denotes the S-dimensional EM estimate of a^j updated from ϕ_n.

$f^j_r(\phi_n) = N^{-1}\sum_{i=1}^N e^j_r(x_i, \phi_n)$ is the frequency of the attributions of x_i to the jth component for the rth drawing of the nth step of the algorithm.

$f^{K+j}_r(\phi_n) = N^{-1}\sum_{i=1}^N e^j_r(x_i, \phi_n) b_j(x_i)$ is an $\mathbb{R}^S$-valued r.v.

$\bar{f}^j_r(\phi_n) = f^j_r(\phi_n) - T^j(\phi_n)$ is the centered r.v. corresponding to $f^j_r(\phi_n)$.

$\bar{f}^{K+j}_r(\phi_n) = f^{K+j}_r(\phi_n) - T^{K+j}(\phi_n)$ is the centered $\mathbb{R}^S$-valued r.v. corresponding to f^{K+j}.

$\Delta^j_r = T^j(\phi_n) f^{K+j}_r(\phi_n) - f^j_r(\phi_n) T^{K+j}(\phi_n)$ is an $\mathbb{R}^S$-valued r.v.

$\Pi^j_n(\phi_n) = T^j(\phi_n)\{\mu_n^{-1}\sum_{r=1}^{\mu_n} f^j_r(\phi_n)\}$ is a real-valued r.v.

Transforming (3.5) and (3.6) by some elementary calculations, we obtain:

Proposition 2. *The sequence ϕ_n generated by MCEM can be expressed as*

$$\phi_{n+1} = T(\phi_n) + \frac{1}{\sqrt{\mu_n}} U(\phi_n, Z_n(\phi_n)) ,$$

where $U(\phi, Z_n(\phi)) = U_n$ is a sequence of $\mathbb{R}^d$-valued r.v.'s such that

$$U_n = (U_n^1, \ldots, U_n^j, \ldots, U_n^K, U_n^{K+1}, \ldots, U_n^{K+j}, \ldots, U_n^{2K}) ,$$

where

$$U_n^j = \frac{\sum_{r=1}^{\mu_n} \bar{f}^j_r(\phi_n)}{\sqrt{\mu_n}}, \quad j = 1, \ldots, K ,$$

is a real-valued r.v., and

$$U_n^{K+j} = \frac{\mu_n^{-1/2} \sum_{r=1}^{\mu_n} \Delta_r^j(\phi_n)}{\Pi_n^j(\phi_n)}, \quad j = 1, \ldots, K ,$$

is an S-dimensional random vector. □

Before stating Theorem 2 we need to introduce the following facts and additional notation:

$\Delta_r^{j,t}$ denotes the tth component of Δ_r^j for $r = 1, \ldots, \mu_n, j = 1, \ldots, K$ and $t = 1, \ldots, S$.

Facts. The centered real-valued r.v.'s $\bar{f}^j_1(\phi), \ldots, \bar{f}^j_{\mu_n}(\phi)$, as well as the centered real-valued r.v.'s $\Delta_1^{j,t}(\phi), \ldots, \Delta_{\mu_n}^{j,t}(\phi)$ are i.i.d. for each $\phi \in G, j = 1, \ldots, K$ and $t = 1, \ldots, S$.

For each integer $p > 0$, the functions $\phi \to E|\bar{f}^j_r(\phi)|^p$ and $\phi \to E|\Delta_r^{j,t}(\phi)|^p$ are continuous on G.

By the S.L.L.N., $\Pi_n^j(\phi)$ converges a.s. to $\Pi^j(\phi) = (T^j(\phi)/N) \sum_{i=1}^N t^j(x_i, \phi)$ for all ϕ in G.

Condition (3.4) ensures that ϕ_n remains in some compact subset G of $\mathbb{R}^d$ and that there exist positive constants A and B such that for all $j = 1, \ldots, K$, ϕ in G and integer $n > 0$,

$$A \leqslant \Pi_n^j(\phi) \leqslant B . \tag{3.7}$$

Theorem 2. *If the operator T satisfies* (C3)–(C4), *the sequence $\gamma_n = \mu_n^{-1/2}$ satisfies* (C8)–(C9) *and* (C10) *and the log-likelihood function L is such that the operator* $D^2L(\phi)$ *is regular then the sequence ϕ_n generated by MCEM converges a.s. to a local maximizer of L, whatever its starting point ϕ_0.*

Proof. We prove that under the assumptions of Theorem 2, we can apply Theorem 1. This enables us to conclude the proof since, from Celeux and Diebolt (1993, Proposition 1), the stable fixed points of T are the proper maximizers of the log-likelihood function. Thus, by Remark 1.1, it is enough to prove the following two points.

(i) The sequence U_n satisfies (C11).

(ii) The sequence U_n satisfies (C12).

Proof of (i). It is enough to establish that each component of U_n satisfies (C11). But, for the k first components U_n^j of U_n, the Chebyshev's and the Dharmadhikari and Jogdao's inequalities (see Dharmadhikari and Jogdao, 1969, or Petrov, 1975, p. 60) imply

$$\sup_{\phi\in G} P\left\{\left|\mu_n^{-1/2}\sum_{r=0}^{\mu_n}\bar{f}^{\,j}_r(\phi)\right|>kr_n\right\}\leqslant\frac{\mathscr{C}(p)}{(kr_n)^p}\leqslant\frac{\mathscr{C}(p)}{k^pn^{p\delta_0}},$$

where $\mathscr{C}(p)$ is a positive constant and p an integer larger than $1/\delta_0$. Moreover, by (3.7) the other components of U_n satisfy

$$P\left\{\left|\Pi_n^j(\phi)^{-1}\mu_n^{-1/2}\sum_{r=0}^{\mu_n}\Delta_r^{j,t}(\phi)\right|>kr_n\right\}\leqslant P\left\{\left|\mu_n^{-1/2}\sum_{r=0}^{\mu_n}\Delta_r^{j,t}(\phi)\right|>kAr_n\right\}.$$

Thus, using the same arguments as above completes the proof.

Proof of (ii). For simplicity's sake, we set throughout this proof $Q_n=\langle u, U(\phi, Z_n(\phi))\rangle_A$. There exists $(\alpha_1,\ldots,\alpha_d)\in\mathbb{R}^d$ such that

$$Q_n=\sum_{j=1}^{K}\alpha_j\sum_{r=1}^{\mu_n}\frac{\bar{f}^{\,j}_r(\phi)}{\sqrt{\mu_n}}+\sum_{j=1}^{K}\sum_{t=1}^{S}\alpha_{K-1+(j-1)S+t}\sum_{r=1}^{\mu_n}\frac{\Delta_r^{j,t}(\phi)}{\Pi_n^j(\phi)\sqrt{\mu_n}}.$$

Let

$$Q_n^1=\sum_{j=1}^{K}\alpha_j\sum_{r=1}^{\mu_n}\frac{\bar{f}^{\,j}_r(\phi)}{\sqrt{\mu_n}}+\sum_{j=1}^{K}\sum_{t=1}^{S}\alpha_{K-1+(j-1)S+t}\sum_{r=1}^{\mu_n}\frac{\Delta_r^{j,t}(\phi)}{\Pi^j(\phi)\sqrt{\mu_n}}$$

and

$$Q_n^2=Q_n-Q_n^1=\sum_{j=1}^{K}\left(\frac{1}{\Pi_n^j(\phi)}-\frac{1}{\Pi^j(\phi)}\right)N_n^j,$$

where

$$N_n^j=\frac{1}{\sqrt{\mu_n}}\sum_{r=1}^{\mu_n}\sum_{t=1}^{S}\alpha_{K-1+(j-1)S+t}\Delta_r^{j,t}(\phi).$$

The proof is organized as follows:

Step 1. We prove that there exist $a_1>0$ and $\rho_1>0$ such that

$$\inf_{\phi\in G}P(Q_n^1>a_1)>\rho_1\quad\text{and}\quad\inf_{\phi\in G}P(Q_n^1<-a_1)>\rho_1.$$

Step 2. We establish that, for any $\delta>0$, $\sup_{\phi\in G}P(|Q_n^2|>\delta)\to 0$ and $n\to\infty$.

Step 3. Using the results of the two previous steps, we achieve the proof.

Proof of Step 1. Denoting

$$W_r(\phi)=\sum_{j=1}^{K}\alpha_{j1}\frac{\bar{f}^{\,j}_r(\phi)}{\sqrt{\mu_n}}+\sum_{j=1}^{K}\sum_{t=1}^{S}\alpha_{K-1+(j-1)S+t}\frac{\Delta_r^{j,t}(\phi)}{\Pi^j(\phi)\sqrt{\mu_n}},$$

Q_n^1 has the form $Q_n^1 = (1/\sqrt{\mu_n})\sum_{r=1}^{\mu_n} W_r(\phi)$, where the r.v.'s $W_r(\phi)$, $r=1,\ldots,\mu_n$, are i.i.d. and nondegenerate, whereas the functions $\phi \to E(|W_r(\phi)|^2)$ and $\phi \to E(|W_r(\phi)|^3)$ are continuous on G. Thus, using Berry–Esseen inequality we obtain the result.

Proof of Step 2. By (3.7), for all positive constants δ and ε we have

$$P(|Q_n^2|>\delta) \leqslant \sum_{j=1}^{K} P\left(|N_n^j| > \frac{\delta}{K\varepsilon}\right) + \sum_{j=1}^{K} P(|\Pi_n^j(\phi) - \Pi^j(\phi)| \geqslant \varepsilon A^2),$$

where

$$\Pi_n^j(\phi) - \Pi^j(\phi) = \frac{T^j(\phi)}{N} \sum_{i=1}^{N} \left\{ \sum_{r=1}^{\mu_n} \frac{(e_r^j(x_i, \phi) - t^j(x_1, \phi))}{\mu_n} \right\}$$

and

$$\sup_{\phi\in G} T^j(\phi) < \infty .$$

The vectors $(e_r^j(x_i),\ r=1,\ldots,\mu_n)$ and $(\mathbf{1}_{[0,t^j(x_i)]}(\boldsymbol{u}_r),\ r=1,\ldots,\mu_n)$ have the same distribution, where $\mathbf{1}_A$ denotes the indicator function of the set A and $\boldsymbol{u}_1,\ldots,\boldsymbol{u}_{\mu_n}$ is a sequence of i.i.d. r.v.'s uniformly distributed on $[0, 1]$. Thus, the Glivenko–Cantelli theorem implies that $\sup_{\phi\in G} P(|\Pi_n^j(\phi) - \Pi^j(\phi)| \geqslant \varepsilon A^2) \to 0$ as $n\to\infty$. Moreover the Berry–Esseen inequality entails that $\sup_{\phi\in G} P(|N_n^j| > \delta/(K\varepsilon)) \to 0$ as $n\to\infty$ and $\varepsilon\to\infty$.

Proof of Step 3. For all $a>0$ and $\delta>0$ we have

$$P(Q_n > a) \geqslant P(Q_n^1 > a+\delta) + P(|Q_n^2| < \delta) - 1 .$$

Provided that δ has been chosen small enough, the Step 1 of the proof entails that there exist $a>0$ and $\rho>0$ such that, $\inf_{\phi\in G} P(Q_n^1 > a+\delta) \geqslant \frac{1}{2}\rho$. Finally, by Step 2 we have $1-\inf_{\phi\in G} P(|Q_n^2|>\delta) < \frac{1}{2}\rho$ for n large enough. We obtain similarly that there exist $a>0$ and $\rho>0$ such that $\inf_{\phi\in G} P(Q_n < -a) > \rho$ for n large enough. □

Appendix

Proposition A. *Let* $0<\rho<1$ *and consider a sequence* j_n *of integers such that* $\sum_{n=0}^{\infty}\rho^{j_n} = \infty$ *and* $\lim_{n\to\infty} j_n = \infty$ *and set* $\tilde{n} = n-1+\sum_{p=1}^{n-1} j_p$. *Let* $(\Omega, \mathscr{F}, P)$ *be a probability space; then denote by* $\{A_k, k\in\mathbb{N}\}$, $\{B_k, k\in\mathbb{N}\}$, $\{C_k, k\in\mathbb{N}\}$ *three sequences of* $\mathscr{F}$ *sets and let* $\{\mathscr{F}_k, k\in\mathbb{N}\}$, $\{\sigma_k, k\in\mathbb{N}\}$ *be two sequences of σ-fields of* $\mathscr{F}$ *sets satisfying the following conditions:*

For all $k \leqslant \tilde{n}+j_n$: $A_k \in \mathscr{F}_n$, $C_k \in \mathscr{F}_n$.

For all $n\in\mathbb{N}$: $A_n \in \sigma_n$, $B_{\tilde{n}} \in \mathscr{F}_{n-1}$ *and* $\sigma_{\tilde{n}+j_n} = \mathscr{F}_n$.

Now, consider the events $E_{n,j_n} = E_{n,j_n}^+ \cup E_{n,j_n}^-$, *where*

$$E_{n,j_n}^+ = B_{\tilde{n}} \bigcap_{p=0}^{j_n} C_{\tilde{n}+p} \quad \textit{and} \quad E_{n,j_n}^- = B_{\tilde{n}}^c \bigcap_{p=0}^{j_n} C_{\tilde{n}+p} .$$

Then, we have the following statement: If

$$E(\mathbf{1}_{A_k}/\sigma_{k-1}) \geqslant \rho \quad a.s. \quad and \quad E(\mathbf{1}_{c_k}/\sigma_{k-1}) \geqslant \rho \quad a.s.$$

for all $k > k_0$, *then*

$$P\left(\limsup_{n\to\infty} E_{n,j_n}\right) = 1 .$$

Proof. By the conditional version of the Borel–Cantelli lemma (see, e.g., Hall and Heyde, 1985, p. 32), it suffices to prove that $\sum_{n=1}^{\infty} P(E_{n,j_n}/\mathscr{F}_{n-1}) = \infty$ a.s. But, since $\mathscr{F}_{n-1} \subset \sigma_{\tilde{n}+j_n-p}$ for $1 \leqslant p \leqslant j_n + 1$, conditioning successively by $\sigma_{\tilde{n}+j_n-1}, \sigma_{\tilde{n}+j_n-2}, \ldots, \sigma_{\tilde{n}-1}$ we obtain

$$P(E^+_{n,j_n}/\mathscr{F}_{n-1}) \geqslant \rho^{j_n+1}\mathbf{1}_{B_{\tilde{n}}} \quad \text{a.s.}$$

and

$$P(E^-_{n,j_n}/\mathscr{F}_{n-1}) \geqslant \rho^{j_n+1}\mathbf{1}_{B^c_{\tilde{n}}} \quad \text{a.s.} ,$$

which implies

$$P(E_{n,j_n}/\mathscr{F}_{n-1}) \geqslant \rho^{j_n+1} \quad \text{a.s.} \qquad \square$$

Corollary A. *Consider the events* $E_{n,j_n} = E^+_{n,j_n} \cup E^-_{n,j_n}$, *where*

$$E^+_{n,j_n} = \{q_{\tilde{n}} \geqslant 0\} \cap \left\{\bigcap_{k=0}^{j_n} \{w_{\tilde{n}+k} > a\}\right\}$$

and

$$E^-_{n,j_n} = \{q_{\tilde{n}} \leqslant 0\} \cap \left\{\bigcap_{k=0}^{j_n} \{w_{\tilde{n}+k} < -a\}\right\} .$$

There exists $a > 0$ *and* $b > 0$ *such that, if* $j_n = [b \log n]$ *then*

$$P\left(\limsup_{n\to\infty} \{E_{n,j_n}\}\right) = 1 .$$

Proof. This result is an immediate consequence of Proposition A. $\square$

Acknowledgement

I wish to thank Professor Jean Diebolt for introducing me to the topic. His personal encouragements and useful criticism were greatly appreciated.

References

J.C. Biscarat, Almost sure convergence of a class of stochastic algorithms, Tech. Rept. No. 155, L.S.T.A. University of Paris VI (Paris, 1992).

J.C. Biscarat, G. Celeux and J. Diebolt, Stochastic versions of the EM algorithm (1992), submitted to: Internat. Statist. Rev.

G. Celeux and J. Diebolt, A stochastic approximation type EM algorithm for the mixture problem, Stoch. Stoch. Rep. 41 (1993) 119–134.

P.G. Ciarlet, Introduction à l'analyse numérique matricielle et à l'optimisation (Masson, Paris, 1985).

A.P. Dempster, N. Laird and D.B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, J. Roy. Statist. Soc. Ser. B 39 (1977) 1–38.

S.W. Dharmadhikari and K. Jogdao, Bounds on moments of certain random variables, Ann. Math. Statist. 40 (4) (1969) 1506–1508.

P. Hall and C.C. Heyde, Martingale Limit Theory and its Application (Academic Press, New York, 1980).

V.V. Petrov, Sum of Independent Random Variables (Springer, Berlin, 1975).

R.A. Redner and H.F. Walker, Mixtures densities, maximum likelihood and the EM algorithm, SIAM Rev. 26 (1984) 195–249.

M.A. Tanner, Tools for statistical inference, Lecture Notes in Statist. No. 67 (Springer, New York, 1991).

D.M. Titterington, A.F.M. Smith and U.E. Makov, Statistical Analysis of Finite Mixture Distribution (Wiley, New York, 1985).

G.C.G. Wei and M.A. Tanner, A Monte Carlo implementation of the EM algorithm and the poor man's data augmentation algorithms, J. Amer. Statist. Assoc. 411 (85) (1991) 699–704.